Journal of Applied Statistics

ISSN: 0266-4763 (Print) 1360-0532 (Online) Journal homepage: https://www.tandfonline.com/loi/cjas20

Rotation of principal components: choice of normalization constraints

Ian T. Jolliffe

Rotation of principal components: choice of normalization constraints

IAN T. JOLLIFFE, *Department of Mathematical Sciences, University of Aberdeen*

SUMMARY *Following a principal component analysis, it is fairly common practice to rotate some of the components, often using orthogonal rotation. It is a frequent misconception that orthogonal rotation will produce rotated components which are pairwise uncorrelated, and/or whose loadings are orthogonal. In fact, it is not possible, using the standard definition of rotation, to preserve both these properties. Which of the two properties is preserved depends on the normalization chosen for the loadings, prior to rotation. The usual 'default' normalization leads to rotated components which possess neither property.*

1 Introduction

Principal component analysis is often used to reduce the dimensionality of a data set, replacing the p variables which have been measured by a much smaller set of m components. To interpret the space spanned by the m retained components, it is fairly common practice to rotate some or all the components, in an attempt to achieve a 'simple structure'. Some drawbacks associated with such rotation have been discussed previously (Jolliffe, 1989). One of these drawbacks is that different choices for the normalization constraints imposed on the loadings in the components will lead to different rotated solutions.

The present paper discusses in detail the choice of normalization constraints. In particular, it is shown that different normalizations have different implications for the dual properties of the orthogonality of rotated loadings and the uncorrelatedness of rotated components. Although both properties are present before rotation, it is impossible to retain the two simultaneously after rotation. The most usual normalization constraint leads to the loss of both properties but, by choosing other normalizations, one of the two properties can be retained.

Section 2 defines three normalization constraints, and Section 3 describes rotation, together with the implications of the three normalization constraints for rotated components. Two examples are presented in Section 4. Discussion of the

0266-4763/95/010029-07

results, including some comments on the differences from factor analysis, is given in Section 5. Derivations of the algebraic results quoted in Section 3 are outlined in the appendix.

2 Normalization constraints

Let us suppose that the kth principal component is denoted by $z_k = \mathbf{a}'_k\mathbf{x}$, where $\mathbf{x}$ is the vector of the original p variables, and $\mathbf{a}_k$ is the vector of coefficients or loadings which define the component. Let us consider the following three possible normalization constraints.

$$\mathbf{a}'_k\mathbf{a}_k = l_k \tag{1}$$

where l_k is the eigenvalue associated with z_k, and is the variance of z_k when $\mathbf{a}'_k\mathbf{a}_k = 1$. With this normalization, the elements of $\mathbf{a}_k$ can be interpreted as correlations between z_k and each of the original variables, assuming (as is often the case) that the variables in $\mathbf{x}$ have been standardized to have unit variance. This normalization is often the default (or only) option in computer packages which carry out rotation. This is probably because of the suitability of the normalization in factor analysis, and most packages which have rotation as an option treat principal component analysis as a 'special case' of factor analysis.

Another normalization constraint is

$$\mathbf{a}'_k\mathbf{a}_k = 1 \tag{2}$$

This normalization arises naturally in the derivation of principal components, and has the property of preserving distances.

A third normalization constraint is

$$\mathbf{a}'_k\mathbf{a}_k = l_k^{-1} \tag{3}$$

With this normalization, the variances of all the components are the same (unity), which is appropriate if it is required that all the components should have equal importance. This normalization is also useful when using principal components to detect outliers (Jolliffe, 1986, Section 10.1).

3 Rotation of principal components

Let us suppose that $\mathbf{X}$ is a $(n \times p)$ matrix whose ith row contains the values of the p original variables for the ith (of n) observations. It is assumed, for convenience, that each column of $\mathbf{X}$ has been centred to have mean zero. Let $\mathbf{Z} = \mathbf{XA}$, where $\mathbf{A}$ is a $(p \times q)$ matrix whose columns are vectors $\mathbf{a}_k$ of loadings associated with a subset of q components which are to be rotated. The ith row of $\mathbf{Z}$ consists of the values of the q selected components for the ith observation.

Rotating the q principal components is achieved by post-multiplying $\mathbf{A}$ by a $(q \times q)$ matrix $\mathbf{T}$ to obtain the rotated loadings $\mathbf{B} = \mathbf{AT}$. Values of the rotated components are elements of the matrix

$$\mathbf{F} = \mathbf{XB} = \mathbf{XAT} = \mathbf{ZT}$$

For orthogonal rotation, $\mathbf{T}$ is an orthogonal matrix. It should be noted that rotation is carried out with respect to the loadings rather than the observations.

For the normalization constraint of equation (1), it is shown in the appendix that

$$\mathbf{B}'\mathbf{B} = \mathbf{T}'\mathbf{L}\mathbf{T}$$

and the covariance matrix for $\mathbf{F}$ is

$$\mathrm{var}(\mathbf{F}) = \mathbf{T}'\mathbf{L}^2\mathbf{T}$$

where $\mathbf{L}$ is a $(q \times q)$ diagonal matrix whose elements are the eigenvalues corresponding to the q selected components. Neither $\mathbf{B}'\mathbf{B}$ nor $\mathrm{var}(\mathbf{F})$ is a diagonal matrix, so the rotated loadings are not orthogonal and the rotated components are not uncorrelated.

For the normalization of equation (2), we have

$$\mathbf{B}'\mathbf{B} = \mathbf{I}_q$$

$$\mathrm{var}(\mathbf{F}) = \mathbf{T}'\mathbf{L}\mathbf{T}$$

where $\mathbf{I}_q$ is the $(q \times q)$ identity matrix. Thus, the rotated loadings are orthogonal but the rotated components are correlated. It should be noted that the optimal rotation matrix $\mathbf{T}$ will generally be different for different normalization constraints. For notational simplicity, $\mathbf{T}$ has been used to denote all such rotation matrices, but the reader should avoid falsely equating apparently identical expressions for different normalizations.

For the normalization of equation (3), we have

$$\mathbf{B}'\mathbf{B} = \mathbf{T}'\mathbf{L}^{-1}\mathbf{T}$$

$$\mathrm{var}(\mathbf{F}) = \mathbf{I}_q$$

Hence, the rotated loadings are not orthogonal but the rotated components are uncorrelated.

4 Examples

The two examples which follow illustrate two extremes of behaviour. It was noted earlier (Jolliffe, 1989) that, if rotation is restricted to components whose corresponding eigenvalues, l_k are all of similar magnitude, then the choice of normalization should have little effect. The second example illustrates this phenomenon. In contrast, if the eigenvalues of the rotated components are widely separated, as is often the case if the first few components are rotated, then different normalizations may produce very different results. This is demonstrated in the first example.

4.1 100 km running data

This data set, which was discussed previously (Jolliffe, 1986, Sections 5.3 and 11.3), consists of the times taken by the 80 finishers for each of ten sections of 10 km in the Lincolnshire 100 km race in June 1984. For a principal component analysis using the correlation matrix, the first three eigenvalues are 7.24, 1.28 and 0.55, accounting for 72%, 13% and 6%, respectively, of the total variation. A routine analysis, as frequently carried out in fields such as psychology and meteorology, would typically decide to retain two components and rotate them to give a simple structure. Table 1 gives the original loadings—with the normalization of equation (2)—for the first two components, together with the rotated loadings corresponding

TABLE 1. Lincolnshire 100 km data: loadings and rotated loadings for three different normalization constraints, and components 1 and 2

Unrotated		Rotated					
		$\mathbf{a}_k'\mathbf{a}_k = l_k$		$\mathbf{a}_k'\mathbf{a}_k = 1$		$\mathbf{a}_k'\mathbf{a}_k = l_k^{-1}$	
Component 1	Component 2	Component 1	Component 2	Component 1	Component 2	Component 1	Component 2
0.30	0.45	0.11	0.46	0.53	−0.07	0.52	−0.24
0.30	0.45	0.12	0.46	0.54	−0.07	0.52	−0.24
0.33	0.34	0.18	0.44	0.47	0.02	0.42	−0.13
0.34	0.20	0.25	0.40	0.38	0.13	0.29	−0.00
0.34	−0.06	0.34	0.28	0.17	0.30	0.04	0.22
0.35	−0.16	0.38	0.25	0.11	0.37	−0.05	0.31
0.31	−0.27	0.39	0.17	−0.00	0.41	−0.17	0.39
0.31	−0.30	0.40	0.16	−0.02	0.43	−0.20	0.42
0.31	−0.29	0.39	0.16	−0.02	0.42	−0.19	0.40
0.27	−0.40	0.39	0.08	−0.13	0.46	−0.30	0.48

TABLE 2. Facial spots data, with outlier removed: loadings and rotated loadings for three different normalization constraints, and components 2 and 3

Unrotated		Rotated					
		$\mathbf{a}_k'\mathbf{a}_k = l_k$		$\mathbf{a}_k'\mathbf{a}_k = 1$		$\mathbf{a}_k'\mathbf{a}_k = l_k^{-1}$	
Component 2	Component 3	Component 2	Component 3	Component 2	Component 3	Component 2	Component 3
0.53	−0.49	0.69	−0.19	0.69	−0.20	−0.20	0.69
−0.03	−0.49	0.19	−0.45	0.19	−0.46	−0.46	0.20
−0.51	0.13	−0.52	−0.12	−0.52	−0.12	−0.11	−0.51
−0.36	0.22	−0.42	0.03	−0.42	0.03	0.04	−0.42
0.57	0.68	0.22	0.86	0.21	0.86	0.86	0.19

to each of the normalizations of equations (1)–(3). For convenience of comparison between the results for the various normalizations, the loadings have been re-normalized after rotation, so that $\mathbf{b}_k'\mathbf{b}_k = 1$, where $\mathbf{b}_k$ is any vector of rotated loadings.

All the rotations are carried out using the varimax criterion, but similar conclusions are reached with other orthogonal rotation criteria. The unrotated loadings are easy to interpret: the first component is a measure of the overall time taken for the race, while the second component contrasts the times taken early in the race with those taken towards the race end. Thus, the second component contrasts those runners who slow down considerably during the race with those whose race is more evenly paced, and is a measure of the degree of slowing down.

With the normalization of equation (1), the rotated components are less easy to interpret: the first component is roughly a measure of the time taken to cover the last 60 km, while the second component is dominated by the time taken to complete the first 40 km. Given that those competitors who run fastest in the first part of the race also tend to run fastest later on, it is intuitively obvious that the two rotated components are correlated. The correlation between them is actually 0.94. In addition, given that all the loadings for both rotated components are positive, it is clear that the vectors of loadings are not orthogonal. In fact, we have $\mathbf{b}_1'\mathbf{b}_2 = 0.70$.

For the normalization of equation (2), the interpretation of the rotated components is essentially the same as for equation (1), apart from an arbitrary reversal of order. The detailed structure is somewhat simpler than that given by equation (1) (fewer intermediate-sized loadings) and, in addition, the two vectors of loadings are orthogonal. However, the rotated components are not uncorrelated—their correlation is 0.70.

Finally, for the normalization of equation (3), the first rotated component is dominated by the time for the first 40 km—as with equation (2)—but it is now contrasted to some extent with the time for the final 40 km. Similarly, the second rotated component is dominated by the time for the last 60 km, but there is a contrast with the first 20 km. The two rotated components are now uncorrelated but the vectors of rotated loadings are not orthogonal, with $\mathbf{b}_1'\mathbf{b}_2 = -0.68$.

In this particular example, it is clear from every point of view (orthogonality, uncorrelatedness and interpretability) that the most usual form of normalization, i.e. equation (1), is not to be recommended. The normalization of equation (2) performs best with respect to interpretability and orthogonality but, if uncorrelatedness is a major concern, then the normalization of equation (3) will be preferred. Of course, as noted at the beginning of the example, it is arguable whether rotation is desirable at all in this case. However, rotation is carried out routinely in some fields of application, and the example illustrates possible problems associated with such routine use.

4.2 Facial spots

These data consist of measurements of the number of facial spots in five categories of severity (i.e. five variables) for 34 individuals, and were discussed previously (Jolliffe, 1989) in the context of principal component rotation. It was shown that, when one outlying observation is removed, the second and third eigenvalues of the correlation matrix are very close (0.74, 0.70). Therefore, the corresponding eigenvectors are ill-defined, and rotation within the subspace defined by these eigenvectors may be useful in interpreting the subspace. Table 2 has the same format as Table 1 but, in this case, is for the second and third eigenvectors in the spots data with one outlier removed.

It was noted at the beginning of this section that, for cases where the eigenvalues of rotated components are similar, then the choice of normalization constraint has little effect. This is illustrated in Table 2—apart from one case where the difference is 0.03, none of the loadings differs by more than 0.01 for the three different normalizations. Furthermore, these small differences ensure that, for all three normalizations, the rotated loadings are (approximately) orthogonal and the rotated components are (approximately) uncorrelated. The largest correlation is 0.04 for the normalization of equation (1), and the largest value of $|\mathbf{b}_2\mathbf{b}_3|$ is 0.02, for the normalizations of equations (1) and (3).

5 Discussion

Previously, it was argued (Jolliffe, 1989) that rotation of principal components, as commonly practiced, has a number of drawbacks, one of which is that different normalizations may lead to qualitatively different results. This phenomenon has been described in detail, and the scale of the possible differences is demonstrated by the first example above. It was also argued (Jolliffe, 1989) that this drawback—and some others—can be avoided if rotation is restricted to sets of components whose corresponding eigenvalues are of similar magnitude. This is illustrated by the second example. Although the strategy of rotating only components with close eigenvalues is a sound strategy, the practice of rotating the first few components, regardless of the spacing of their eigenvalues, is useful for some purposes and is likely to continue. It is desirable, therefore, that such rotations are not carried blindly, without consideration of the choice of normalization constraint and of the implications of the choice for the properties of the rotated components and their loadings.

The results presented above are not widely appreciated. For example, the *SAS/STAT User's Guide* (SAS, 1990, p. 786), in describing rotation within the FACTOR procedure (rotation of principal components must be done using this procedure; *SAS/STAT User's Guide*, p. 1249), states that, "if the factors are rotated by an orthogonal transformation, the rotated factors are also uncorrelated". Similarly, Dillon and Goldstein (1984, p. 91) stated that "Factors resulting from the orthogonal rotation of principal components will remain statistically uncorrelated: that is, the cosine of the angle between rotated factors is zero". What neither of these sources—nor a number of others—makes clear is that the uncorrelatedness property only holds for one particular, rarely used normalization. The latter part of the second quotation adds extra confusion—are the authors discussing rotated loadings or rotated scores?

Rencher (1992) made the opposite statement to those cited above, commenting that, "... rotated components are no longer uncorrelated ...". However, he did not note that his statement is specific to one particular normalization. In terms of specific applications, a recent example where an apparently incorrect statement is made is in an analysis of thunderstorm activity in the south-eastern USA by Easterling (1991), who stated that, "Orthogonal rotations maintain the constraint that eigenvectors must be temporally uncorrelated". Other examples exist, especially in the climatological literature, but often the incorrect assumptions are implicit rather than explicit.

Finally, it should be stressed that the discussion above refers to rotated principal components and not to factor analysis. In factor analysis, factor scores are not computed directly using rotated factor loadings, but are estimated using an indirect procedure. Such a procedure may produce uncorrelated rotated factor scores, even with the normalization of equation (1). In this case, a factor should not be interpreted in terms of the variables by means of the rotated loadings, but that is a separate concern.

Correspondence: I. T. Jolliffe, Department of Mathematical Sciences, University of Aberdeen, Edward Wright Building, Dunbar Street, Aberdeen AB9 2TY, UK.

REFERENCES

DILLON, W. R. & GOLDSTEIN, M. (1984) *Multivariate Analysis Methods and Applications* (New York, Wiley).

EASTERLING, D. R. (1991) Climatological patterns of thunderstorm activity in South-Eastern USA, *International Journal of Climatology*, 11, pp. 213–221.

JOLLIFFE, I. T. (1986) *Principal Component Analysis* (New York, Springer).

JOLLIFFE, I. T. (1989) Rotation of ill-defined principal components, *Applied Statistics*, 38, pp. 139–147.

RENCHER, A. C. (1992) Interpretation of canonical discriminant functions, canonical variates, and principal components, *American Statistician*, 46, pp. 217–225.

SAS INSTITUTE (1990) *SAS/STAT User's Guide*, Version 6, 4th edn (Cary, NC, SAS Institute).

Appendix: Derivation of results in Section 3

With the notation in Section 3, we have $\mathbf{Z}=\mathbf{XA}$, $\mathbf{B}=\mathbf{AT}$ and $\mathbf{F}=\mathbf{XB}$. The matrix $\mathbf{B}$ has columns which are vectors $\mathbf{b}_k$ of loadings for the rotated components.

$$\mathbf{B'B}=\mathbf{T'A'AT}$$
$$=\mathbf{T'LT}$$

for the normalization of equation (1), where $\mathbf{L}$ has kth diagonal element $l_k=\mathbf{a}_k'\mathbf{a}_k$.

For the normalization of equation (2), $\mathbf{A'A}=\mathbf{I}_q$, because $\mathbf{a}_k'\mathbf{a}_k=1$, so $\mathbf{B'B}=\mathbf{T'T}=\mathbf{I}_q$, since $\mathbf{T}$ is orthogonal. Similarly, for the normalization of equation (3), $\mathbf{A'A}=\mathbf{L}^{-1}$, and $\mathbf{B'B}=\mathbf{T'L}^{-1}\mathbf{T}$.

Turning to $\mathbf{F}$, we have, for example, that

$$\mathrm{var}(\mathbf{F})=\mathbf{B}'\,\mathrm{var}(\mathbf{X})\mathbf{B}=\mathbf{B'SB}$$
$$=\mathbf{T'A'SAT}$$

The columns of $\mathbf{A}$ are eigenvectors of $\mathbf{S}$, and, with the normalization of equation (2), we have $\mathbf{A'SA}=\mathbf{L}$, so that

$$\mathrm{var}(\mathbf{F})=\mathbf{T'LT}$$

For the normalizations of equations (1) and (3), it follows that $\mathbf{A'SA}=\mathbf{L}^2$, $\mathbf{A'SA}=\mathbf{I}_q$, respectively, so that $\mathrm{var}(\mathbf{F})=\mathbf{T'L}^2\mathbf{T}$ and $\mathrm{var}(\mathbf{F})=\mathbf{T'T}=\mathbf{I}_q$ in these two cases.